Trip Report – OECD-NEA-WPNCS Expert Group Meetings
Paris, France, 2-6 July 2018

Forrest B. Brown, Michael E. Rising
Monte Carlo Methods, Codes. & Applications Group (XCP-3)
Los Alamos National Laboratory, PO Box 1663, MS A143, Los Alamos NM 87545

Trip Report Date: 2018-09-10, Unclassified (DC: FBB)

1.0 Introduction

During 2-6 July 2018, Brown and Rising participated in the OECD-NEA-WPNCS Expert Group meetings on Advanced Monte Carlo Techniques (EGAMCT), Uncertainty Analysis for Criticality Safety Applications (EGUACSA), and various Subgroups of the WPNCS. It should be noted that LANL has been an active participant in these international meetings for over 15 years, and that there have been very significant benefits from that participation. International collaboration is a very effective means of peer review for new ideas and approaches to computational methods. The interchange of new ideas both for what works and for what doesn't work can save substantial amounts of development time, lead to better understanding of novel approaches, and improve the quality of advanced methods. These benefits are especially true for the OECD-NEA-WPNCS meetings due to a common focus on implementing new methods that benefit the end users – nuclear criticality safety practitioners.

The Working Party for Nuclear Criticality Safety (WPNCS) is in the final stages of a reorganization of the expert group structure, with 2018 the last year for any expert groups. Going forward, the work of the expert groups will be handled by "Subgroups" of the WPNCS. Subgroups will focus on a specific topic with one or more deliverables and are intended to exist for only 2 years or less. The perceived advantages of this new approach are that subgroups will be narrowly focused, with specific deliverables and short life, making it easier for the WPNCS to manage and assess progress. The actual disadvantages are that there is essentially no WPNCS coordination or planning of the different subgroups, the short life precludes anything except simple well-defined work tasks, and the subgroup naming (currently SG-1 through SG-6) makes it nearly impossible to understand what work is in progress.

Brown and Rising participated in EGAMCT, EGUACSA, SG-1, SG-2, SG-3, SG-4, and the WPNCS executive meeting. (Did not attend SG-5.) Summaries of these activities are presented in the sections that follow.

2.0 EGAMCT meeting

About 25 people were present at this meeting, from the US, France, Japan, UK, Germany, Sweden, and a few other places. Shuichi Tsuda (NEA) began with a summary of the agenda and minutes from the 2017 meeting.

Discussions on Final Report

Eric Dumonteil (IRSN) is the current chair of the EG and summarized the work currently in progress. Computations for benchmark problems and analysis tasks are finished, and the focus of the group is to

complete the final report by the end of 2018. We spent some time discussing the draft report, with at least some comment from each of the participants.

The final report addresses the issue of undersampling in Monte Carlo (MC) calculations, where some local tallies (e.g., detectors, flux-wires, fission distribution, etc.) are severely underestimated due to a lack of neutron coverage of the tally regions. The severe case of undersampling is called clustering. At the time of the meeting, none of the participants had been able to devise a reliable, robust method for detecting the presence of undersampling or clustering. All agreed that such a diagnostic test is badly needed, and is a necessary first step toward resolving the undersampling problem.

Brown mentioned a few specific items regarding the draft report: First, it is important to clearly explain that the clustering observed in some of the computations is a calculational artifact of the iteration scheme used in Monte Carlo k-effective calculations; it is distinct from clustering that has been observed in physical systems with low populations subject to stochastic transport effects. Second, the draft report suggests that 1,000 or more neutrons/cycle be used in k-effective calculations. LANL experience has led to a significantly larger recommendation – 10,000 neutrons/cycle for routine NCS problems, or 100,000 neutrons/cycle or more for reactors, large solution tanks, and loosely-coupled problems. Third, the draft report discusses the well-known underestimation of confidence intervals for local tallies in criticality problems, but did not mention the half-dozen or so approximate correction methods that have been tried in the past to produce more-correct confidence intervals. Brown agreed to contribute some paragraphs and references to the draft report on these items.

Brown also briefly mentioned a possible recent breakthrough for diagnosing undersampling in criticality problems. Because the diagnostic was brand new and being investigated (i.e., did not exist until after the EG had completed its work), it would not be included in the final report. Further study and assessment is needed. This work forms the basis for a new Subgroup (discussed in following sections).

Mennerdahl (Sweden) raised objections to most parts of the draft report, but provided no coherent explanation or justification for his objections. He cited his own Excel spreadsheet work (which did not work during the 2017 meeting), and was not supportive of any of the theoretical foundations supporting the EGAMCT work. The rest of the group urged him to produce a written summary of his concerns, to be perhaps included as an appendix to the final report.

Technical Presentations

Brown made a presentation on the release of MCNP6.2 and Whisper, including some brief verification/validation results. (Rising made the same presentation at the DOE-NCSP Technical Program Review in March 2018.)

Dumonteil made a presentation on preliminary results of the neutron clustering experiments performed at RPI. (Same as March 2018 presentation at the DOE-NCSP TPR.)

Andrea Zoia (CEA) made a presentation on neutron transport in random media. His studies involved random tessellations of simple spherical or box geometries, either embedded in fuel-pin lattices or comprising the entire geometry. The random geometries were created by a standalone code, then used as input to repeated Tripoli replica calculations. He looked at average chord lengths for the random and unperturbed problems. While this work is potentially important for analyzing damaged fuel following reactor accidents, none of the work was actually new. All of the material presented had been intensely analyzed in the 1960-70s (at the US naval nuclear laboratories); Zoia seemed unaware of that work, even though it was published in publicly-accessible reports and journals. (Brown had done further work on the topic in the 1980s.)

Kenneth Burns (Italy, ENEA) made a presentation on hybrid methods for coupling MC criticality calculations (that traditionally do not involve variance reduction) with external detector response calculations (that do require extensive variance reduction). While this work is important, it is very much still in the R&D phase. We will of course keep watch on this research as it progresses.

Discussion of Follow-on Subgroups

With the demise of EGAMCT at the end of 2018, there was discussion of proposals for Subgroups to continue and extend the EGAMCT work. Brown and Zoia each presented 2 proposals.

Brown's proposals involved statistical testing to be applied to 2 distinct portions of MC criticality calculations. First, it is possible to perform statistical tests to determine convergence of the source iteration scheme to the fundamental mode solution. At the time, Brown was investigating 6 different statistical tests for that purpose. The Subgroup proposal was to involve others in the NCS community to assess the 6 tests for convergence and possibly propose additional ones. Second, it is possible to detect undersampling (after convergence) using 2 statistical tests involving Shannon entropy. The tests are based on and consistent with the undersampling behavior observed in the EGAMCT work. The Subgroup proposal was to involve others in the NCS community to assess the 2 tests for undersampling and possibly propose additional tests. After much discussion, it was decided to combine the 2 proposals into one Subgroup that would study statistical testing for convergence and undersampling. Brown combined the proposals and presented the agreed-upon proposal to the WPNCS later in the week. The WPNCS agreed to accept the proposal, and the results is a new WPNCS Subgroup: **Subgroup-6, "Statistical tests for diagnosing fission source convergence and undersampling in Monte Carlo criticality calculations."** The final, accepted Subgroup mandate is attached to this report.

Zoia's 2 proposals involved studying geometric perturbations to MC criticality calculations and the impact of clustering on MC depletion calculations. After much discussion, it was agreed that his first proposal would be narrowed in scope and presented to WPNCS in 2019 for a new subgroup. The other proposal concerning clustering and depletion would be tabled for now, and perhaps considered in the future.

3.0 Subgroup-1, Role of Integrated Experiment Uncertainties and Covariance Data in Criticality Safety Validation

About 15 people attended this initial meeting of Subgroup-1 (SG-1). It essentially is a continuation of the previous EGUACSA Phase-IV benchmark. The meeting was opened by the coordinator Shuichi Tsuda (NEA) with a welcome and a review of the agenda and deliverables of the newly formed SG-1.

Discussions on Draft Report

Maik Stuke (GRS) opened the meeting with a discussion of the current draft of the "State-of-the-Art" report. Because this SG-1 is a continuation of the EGUACSA Phase IV study group, many of the results have already been collected. However, as was the case with the previous meeting, some of the results from some of the participants within the current report are difficult to completely comprehend. It seemed that the status of the report remains temporarily on hold due to some peculiar results from one of the participants, Mennerdahl (Sweden). While it is important to include all participants' results in the final report, the consensus was that they would like to understand the source of the peculiar results and address these discrepancies in the write-up prior to finalizing the work. During this discussion of the final report, with respect to the peculiar results, Mennerdahl conceded that the complexity of the problem might have

led to a misunderstanding which is the cause for the peculiar results. This still needs to be documented in the final "State-of-the-Art" report.

One item mentioned in the discussion of the final report was the possibility of any follow-on subgroups to continue aspects of this work. Because of the complexity of some of the study cases, it was made clear that any follow-on subgroups need to establish a well-defined benchmark case or two and not allow the scope of the study group to expand. If a particular aspect of the new subgroup became interesting, it could ultimately form the basis for an additional follow-on subgroup in the future. This will allow for a timely final report with a specific goal and message from the outset of the subgroup.

Technical Presentations

Fabian Sommer (GRS) presented on the calculated benchmark correlation coefficients of LCT-097 within a collaboration effort between GRS, SNL and ORNL. The results appeared similar to those presented previously by Marshall (ORNL) with additional correlations estimated.

B.J. Marshall (ORNL) presented on the calculated benchmark correlation coefficients of HST-001. During the presentation, some nomenclature was clarified regarding the definition of correlation coefficients. A consensus was made among the group regarding the nomenclature. Additionally, the "high-fidelity" results were compared to those "low-fidelity" correlations estimated by Ian Hill (NEA). His presence for the remainder of the SG-1 meeting was requested.

Ian Hill (NEA) opened a discussion (which was not originally included on the agenda) on the "low-fidelity" correlations that he established in a spreadsheet intended for the DICE database. This unplanned discussion was helpful in clarifying the methods and results from the "low-fidelity" benchmark correlations.

Axel Hoefer (AREVA) presented on the use of benchmark correlations in licensing applications. He outlined the details of a variation on another statistical method which takes benchmark correlations into account during the safety margin calculations.

Paul Smith (UK, AmecFosterWheeler) presented on a variety of validation methods that the MONK code validation package has implemented over the years. This includes methods like GLLSM, MOCABA (Hoefer), and others. It was also mentioned that they would like to try an extreme value theory-based method such as the one in the Whisper code from LANL.

4.0 Subgroup-2, Blind Benchmark on MOX Damp Powders

About 24 people attend this first meeting of Subgroup-2 (SG-2). It essentially is a continuation of the previous EGUACSA Phase V benchmark. The meeting was opened by the coordinator Shuichi Tsuda (NEA) with a welcome and a review of the agenda and deliverables of the newly formed SG-2 by the chair Coralie Carmouze (CEA).

Discussions on Current Results and Preliminary Draft Report

Coralie Carmouze (CEA) opened with a discussion on the goals of this subgroup, with the primary intention to compare various data adjustment methods, like GLLSM, between different codes and institutions. Much like the EGUACSA Phase IV (now SG-1), many of the results have already been collected. Most of the collected results came from the presentations made in 2017 including: Carmouze (CEA) using Tripoli, Nicolas Leclaire (IRSN) using Moret5 and Tsunami/Tsurfer, and Chris Perfetti

using Scale/Keno/Tsunami/Tsurfer along with a standalone extreme value theory implementation (similar to Whisper).

All of the currently collected results were presented after the technical presentations were made. Many of the results were very easily explained because there was general consistency among results obtained using the same nuclear data and/or covariance data regardless of the code used. All of the results were actually based on the GLLS method because it produces all of the quantities required of the participants in the study. Exploring different techniques to do this data adjustment might be a worthwhile follow-on subgroup. The idea of follow-on subgroups was discussed, but no concrete plans or directions were proposed.

Technical Presentations

Dennis Mennerdahl (Sweden) presented results on using Tsurfer (GLLSM) with the Scale code system to perform the tasks within the SG-2 blind benchmark study. The benchmarks selected for the Tsurfer analysis seemingly came from expert judgment. This was noted during discussion, as well as the fact that very few benchmarks were selected overall and even fewer had meaningful correlations with the blind benchmark applications being studied. Because of the limitations within the selected benchmarks, the results were deemed to be suspect.

Michael Rising (LANL) presented on a variety of methods explored recently to predict the bias for these blind benchmark cases being studied. For all of the benchmark cases, three methods were presented: 1) extreme value theory (EVT) method within Whisper, 2) GLLS method within Whisper using all benchmarks, and 3) a new Machine Learning (ML) approach also applied across the entire suite of Whisper benchmarks. Because the details of the EVT and GLLS methods are rather well known, a bit more detail on the ML algorithms used, namely decision trees, was discussed. Using the benchmark sensitivity profiles as the features in the ML algorithms, it was shown that the bias could be predicted reasonably well across the suite of Whisper benchmarks as well as for the blind benchmark cases. In all cases the EVT method was most conservative with the largest predicted bias and generally the ML method predicted the smallest bias. Some of this work was provided by a LANL summer student (Pavel Grechanuk, OSU) during the summer of 2017 and we will contribute results to the SG-2 chair by the end of calendar year 2018.

5.0 Subgroup-3, A Benchmark Examining the Effect of Temperature on the Neutron Multiplication Factor for PWR Fuel Assemblies

About 25 people attended this initial meeting of Subgroup-3. This Subgroup task was motivated by the changes in UK requirements for transportation of fissile materials. Criticality calculations of a PWR-type fuel assembly are to be performed at 253 K, 293 K, 333 K, and 588 K for 3 burnup cases (fresh, 30 GWd/T, 45 GWd/T). The models should include 13 actinides and 15 fission products. Densities and dimensions are held constant for all temperatures. For the calculations, the nuclide number densities are specified for all cases (i.e., burnup calculations are not done). Participants are to use whatever methods and nuclear data they have available for determining k-effective at each of the temperatures. The object of this Subgroup is to compare results from different sites. NCS practitioners traditionally do nearly all calculations at room temperature, and are not used to handling temperature effects in their codes or nuclear data. Doppler broadening of the nuclear data is expected to significantly affect the calculational results.

6.0 Subgroup-4, Analysis of Past Criticality Accident

About 26 people attended this initial meeting of Subgroup-4. The participants discussed a number of past criticality accidents, with the goal of choosing one for detailed analysis using modern codes, nuclear data, and feedback mechanisms. They are interested in getting better estimates of what happened, not just conservative estimates. Among the past accidents discussed: Windscale 1970, JCO criticality accident, Y12 1958, and others. It was not clear whether any decision was made to focus on one particular past accident.

7.0 Subgroup-5, Experimental Needs for Criticality Safety Purposes

Could not attend.

Appendix A – New Subgroup for WPNCS

Proposal for a New Sub-Group for the OECD-NEA-WPNCS
Forrest Brown, LANL, 2018-07-04

Over the past several years, the OECD-NEA-WPNCS Expert Group on Advanced Monte Carlo Techniques investigated the phenomena of clustering and undersampling in Monte Carlo criticality calculations. A previous Expert Group on Monte Carlo Source Convergence developed breakthrough methods for graphically assessing the initial convergence of the Monte Carlo fission source, using Shannon entropy or Brownian bridge metrics. Much was accomplished in understanding these phenomena, from both theoretical and practical approaches. Those efforts have led to a number of ideas and challenges for new subgroup study topics summarized below.

SG-6 – Statistical tests for diagnosing fission source convergence and undersampling in Monte Carlo criticality calculations

There is a very strong need for statistical testing to determine fission source convergence in Monte Carlo criticality calculations. Automation of such tests will greatly streamline and support the work carried out by NCS practitioners. Recent R&D work has shown that no single statistical test for convergence is sufficiently reliable, robust, and "guaranteed." However, a combination of several standard statistical tests for the similarity of distributions, coupled with a high-fidelity estimate of the fission-matrix source is sufficiently robust, reliable, and repeatable that convergence can be "guaranteed."

During the course of the EG-AMCT studies, a number of statistical metrics and tests were proposed for diagnosing clustering and undersampling. None of these was robust and reliable enough for practical use in production codes. However, the expert group efforts came close. Some recent R&D work stemming from those past efforts has been very successful and promising.

This Sub-Group will provide international input and collaboration on the development and implementation of statistical tests for convergence, with the primary goal of having the MC codes automatically detect convergence (or lack thereof). Newly proposed statistical tests to detect undersampling (after convergence) will also be reviewed.

It should take about 1 year to investigate and assess this approach and a 2nd year to implement in codes and write a report. Note that additional investigation may lead to an even more robust & reliable test.

WPNCS Activity Proposal Sheet – Subgroup-6	
Activity Title	Statistical tests for diagnosing fission source convergence and undersampling in Monte Carlo criticality calculations
Duration	2 Years
Objective & Scope	Assess combinations of statistical metrics and tests for diagnosing convergence of the fission source in Monte Carlo criticality calculations. Also assess newly proposed statistical tests for diagnosing the undesired effects of clustering and undersampling (after convergence). Recommend a robust, reliable, quantitative combination of tests for convergence and undersampling.
Deliverables	Final technical report
Technical significance and priority in the criticality safety evaluation	There is a very strong need for statistical testing to determine fission source convergence in Monte Carlo criticality calculations. Automation of such tests will greatly streamline and support the work carried out by NCS practitioners. Recent R&D work has shown that no single statistical test for convergence is sufficiently reliable, robust, and "guaranteed." However, a combination of several standard statistical tests for the similarity of distributions, coupled with a high-fidelity estimate of the fission-matrix source is sufficiently robust, reliable, and repeatable that convergence can be "guaranteed." Recent work and the final report from the EG-AMCT demonstrated the existence of clustering in the fission source distribution in Monte Carlo criticality calculations. The use of more neutrons/batch reduces or eliminates the clustering, but the central question of diagnosing the problem remains. This problem may occur in analysing criticality for large solution tanks and reactors, and could lead to incorrect results. A robust & reliable statistical test for clustering is badly needed.
Milestones (Timescale)	Start: Accepted proposal and establishment of Subgroup 2 months: Initial email to participants regarding recent R&D and promising statistical metrics & tests 6 months: Discussion by email. First results received & suggestions made for variations, alternate methods, or improvements. 1 year: At OECD-NEA-WPNCS meetings: Discussion of results, discussion of robustness & reliability of tests, planning for final report 1.5 year: Review status of final report 2 years: Report complete and available for final review at WPNCS meeting
Lead organization and co-ordination	Forrest Brown (LANL, USA)
Monitor(s)	John Bess (INL, USA)
Participants (Individuals and organizations)	Michael Rising (LANL, USA) Christopher Perfetti (UNM, USA) Andrea Zoia (CEA, FR) Eric Dumonteil (IRSN, FR) Paul Smith (WOOD, UK) Brad Rearden (ORNL, USA) Fabian Sommes (GRS, DE) Brian Kiedrowski (Univ. Michigan) Colin Josey (LANL, USA) William Martin (Univ. Michigan) Further participants likely once proposal is circulated.